**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
**C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00**

**A2**

(11) International Publication Number: **WO 00/58473**

(43) International Publication Date: 5 October 2000 (05.10.00)

(21) International Application Number: PCT/US00/08621

(22) International Filing Date: 31 March 2000 (31.03.00)

(30) Priority Data:

| | | |
|---|---|---|
| 60/127,607 | 31 March 1999 (31.03.99) | US |
| 60/127,636 | 2 April 1999 (02.04.99) | US |
| 60/127,728 | 5 April 1999 (05.04.99) | US |
| 09/540,763 | 30 March 2000 (30.03.00) | US |

(63) Related by Continuation (CON) or Continuation-In-Part (CIP) to Earlier Applications

| | |
|---|---|
| US | 60/127,607 (CIP) |
| Filed on | 31 March 1999 (31.03.99) |
| US | 60/127,636 (CIP) |
| Filed on | 2 April 1999 (02.04.99) |
| US | 60/127,728 (CIP) |
| Filed on | 5 April 1999 (05.04.99) |
| US | 09/540,763 (CIP) |
| Filed on | 30 March 2000 (30.03.00) |

(71) Applicant *(for all designated States except US)*: CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven, CT 06516 (US). LEACH, Martin [GB/US]; 884 School Street, Webster, MA 01570 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**

*Without international search report and to be republished upon receipt of that report.*

(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"

(57) Abstract

The present invention provides open reading frames ORFX, encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.

# NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

## BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

## SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "ORFX".

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-3161), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1-3161. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an ORFX polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

In a another aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample.

The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

## DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are 3161 novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptide is referred to as a "ORFX polypeptide" or ORFX protein". For example, an ORFX nucleic acid according to the invention is a nucleic acid including an ORF1 nucleic acid, and an ORF polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an ORF1 polypeptide. Unless indicated otherwise, "ORFX" is meant to refer to any of the ORF1-3161 sequences disclosed herein.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides are summarized in Table 1. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention is provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 1 of Table 1, entitled "ORF #", denotes an ORF number assigned to a nucleic acid containing an open reading frame according to the invention.

Column 2 of Table 1, entitled "Internal Identification number (Nucleic Acid Sequence Identification Number, Polypeptide Sequence Identification Number), provides an internal identification number for the indicated ORF, along with sequence identification numbers (SEQ ID NOs.) corresponding to the indicated ORF. In general, for an ORF*n* according to the invention (wherein *n* is any integer from 1 to 3161), a nucleic acid corresponding to the ORF is SEQ ID NO:2*n*-1, and an amino acid sequence encoded by the ORF is SEQ ID NO:2*n*. For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO:1, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO:2. Similarly, a

nucleic acid sequence corresponding to an ORF4 nucleic acid is SEQ ID NO:7, and a polypeptide sequence corresponding to an ORF4 polypeptide is SEQ ID NO:8; a nucleic acid sequence corresponding to an ORF198 nucleic acid sequence is SEQ ID NO:395, and a polypeptide sequence corresponding to an ORF198 polypeptide is SEQ ID NO:396. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention are provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 2 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. These can be retrieved from http://www.ncbi.nlm.nih.gov/.

To determine similarity to previously described proteins, polypeptides encoded by ORFX DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected. Open reading frames were extended beyond the region of the protein matched using standard DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 3 of Table 3, entitled "Protein Domains", lists previously described protein domains, designated by pfam entries, that are present in polypeptides encoded by the ORFs, Also included in column 3 are proteins in which these domains are present. The pfam entries can be retrieved from http://pfam.wustl.edu/. DNA sequences were translated in all six frames and tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at http://pfam.wustl.edu). Translated DNA sequences that matched a protein domain entry in the Pfam database AND had a score of '7.5' were selected.

Column 4 of Table 3, entitled "Protein Classification", lists the type of classification assigned for the protein, based on its homology. Examples of proteins in the classification include the following proteins:

**What is claimed is:**

1. An isolated nucleic acid molecule encoding a polypeptide comprising an amin acid sequence that is at least 85% identical to a polypeptide including an amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is any integer 1-3161, or the complement thereof.

2. The isolated nucleic acid molecule of claim 1, said molecule hybridizing under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule comprising the sequence of nucleotides selected from the group consisting of SEQ ID NO:2*n*- wherein *n* is any integer 1-3161, or the complement thereof.

3. The isolated nucleic acid molecule of claim 1, said molecule encoding a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ I NO: 2*n*, wherein n is any integer 1-3161, or an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of S ID NO: 2*n*.

4. The isolated nucleic acid molecule of claim 1, wherein said molecule encodes polypeptide comprising the amino acid sequence selected from the group consisting of SEQ I NO: 2*n*, wherein *n* is any integer 1-3161.

5. The isolated nucleic acid molecule of claim 1, wherein said molecule comprise the sequence of nucleotides selected from the group consisting of SEQ ID NO:2*n*-1, wherein any integer 1-3161, or the complement thereof.

6. An oligonucleotide less than 100 nucleotides in length and comprising at least contiguous nucleotides selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is a integer 1-3161, or the complement thereof.

7. A vector comprising the nucleic acid molecule of claim 1.

8. The vector of claim 7, wherein said vector is an expression vector.

9 A host cell comprising the isolated nucleic acid molecule of claim 1.

10. A substantially purified polypeptide comprising an amino acid sequence at least 80% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2*n*, wherein *n* is any integer 1-3161.

11. The polypeptide of claim 10, wherein said polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 2*n*, wherein *n* is any integer 1-3161.

12. An antibody that selectively binds to the polypeptide of claim 10.

13. A pharmaceutical composition comprising a therapeutically or prophylactically effective amount of a therapeutic selected from the group consisting of:

a) the nucleic acid of claim 1;

b) the polypeptide of claim 10; and

c) the antibody of claim 12;

and a pharmaceutically acceptable carrier.

14. A kit comprising in one or more containers, a therapeutically or prophylactically effective amount of the pharmaceutical composition of claim 13.

15. A method of producing the polypeptide of claim 10, said method comprising culturing the host cell of claim 9 under conditions in which the nucleic acid molecule is expressed.

16. A method of detecting the presence of the polypeptide of claim 10 in a sample, comprising contacting the sample with a compound that selectively binds to said polypeptide under conditions allowing the formation of a complex between said polypeptide and said

compound. and detecting said complex, if present, thereby identifying said polypeptide in said sample.

17. A method of detecting the presence of a nucleic acid molecule of claim 1 in a sample, the method comprising contacting the sample with a nucleic acid probe or primer that selectively binds to the nucleic acid molecule and determining whether the nucleic acid probe c primer bound to the nucleic acid molecule of claim 1 is present in the sample.

18. A method for modulating the activity of the polypeptide of claim 10, the methoc comprising contacting a cell sample comprising the polypeptide of claim 10 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptid

19. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a ORFX-associated disorder, wherein said therapeutic is selected fro the group consisting of:

a) the nucleic acid of claim 1;

b) the polypeptide of claim 10; and

c) the antibody of claim 12.

20. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) contacting a test compound with the polypeptide of claim 10; and

b) determining if said test compound binds to said polypeptide,

wherein binding of said test compound to said polypeptide indicates the test compound is a modulator of activity or of latency or predisposition to an ORFX-associated disorder.

21. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) administering a test compound to a test subject at an increased risk ORFX-associated disorder, wherein said test subject recombinantly expresses a polypeptide encoded by the nucleotide of claim 1;

compound, and detecting said complex, if present, thereby identifying said polypeptide in said sample.

17. A method of detecting the presence of a nucleic acid molecule of claim 1 in a sample, the method comprising contacting the sample with a nucleic acid probe or primer that selectively binds to the nucleic acid molecule and determining whether the nucleic acid probe c primer bound to the nucleic acid molecule of claim 1 is present in the sample.

18. A method for modulating the activity of the polypeptide of claim 10, the methoc comprising contacting a cell sample comprising the polypeptide of claim 10 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptid

19. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a ORFX-associated disorder, wherein said therapeutic is selected fro the group consisting of:

a) the nucleic acid of claim 1;

b) the polypeptide of claim 10; and

c) the antibody of claim 12.

20. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) contacting a test compound with the polypeptide of claim 10; and

b) determining if said test compound binds to said polypeptide,

wherein binding of said test compound to said polypeptide indicates the test compound is a modulator of activity or of latency or predisposition to an ORFX-associated disorder.

21. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) administering a test compound to a test subject at an increased risk ORFX-associated disorder, wherein said test subject recombinantly expresses a polypeptide encoded by the nucleotide of claim 1;

<212> DNA
<213> Homo sapiens

<400> 3239
aaaaccaaag attctcctgg agttttctct aaactgggtg ttctcctgag gagagtgaca
60
agaaacttgg tgagaaataa gctggcagtg attacgcgtc tccttcagaa tctgatcatg
120
ggtttgttcc tccttttctt cgttctgcgg gtccgaagca atgtgctaaa gggtgctatc
180
caggaccgcg taggtctcct ttaccagttt gtgggcgcca ccccgtacac aggcatgctg
240
aacgctgtga atctgtttcc cgtgctgcga gctgtcagcg accaggagag tcaggacggc
300
ctctaccaga agtggcagat gatgctggcc tatgcactgc acgtcctccc cttcagcgtt
360
gttgccacca tgattttcag cagtgtgtgc tactggacgc tgggcttaca tcctgaggtt
420
gcccgattgg gt
432

<210> 3240
<211> 144
<212> PRT
<213> Homo sapiens

<400> 3240
Lys Thr Lys Asp Ser Pro Gly Val Phe Ser Lys Leu Gly Val Leu Leu
1 5 10 15
Arg Arg Val Thr Arg Asn Leu Val Arg Asn Lys Leu Ala Val Ile Thr
20 25 30
Arg Leu Leu Gln Asn Leu Ile Met Gly Leu Phe Leu Leu Phe Phe Val
35 40 45
Leu Arg Val Arg Ser Asn Val Leu Lys Gly Ala Ile Gln Asp Arg Val
50 55 60
Gly Leu Leu Tyr Gln Phe Val Gly Ala Thr Pro Tyr Thr Gly Met Leu
65 70 75 80
Asn Ala Val Asn Leu Phe Pro Val Leu Arg Ala Val Ser Asp Gln Glu
85 90 95
Ser Gln Asp Gly Leu Tyr Gln Lys Trp Gln Met Met Leu Ala Tyr Ala
100 105 110
Leu His Val Leu Pro Phe Ser Val Val Ala Thr Met Ile Phe Ser Ser
115 120 125
Val Cys Tyr Trp Thr Leu Gly Leu His Pro Glu Val Ala Arg Leu Gly
130 135 140

<210> 3241
<211> 492
<212> DNA
<213> Homo sapiens

<400> 3241
gtggaatttt tttagacaaa gtctcaaaaa acaaacaaac aaacaaaagg taagataaat
60